



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/243737/>

Version: Accepted Version

Article:

Wang, Z., Shi, L., Xu, S. et al. (2026) Memory-augmented autoencoding with self-supervised learning for unsupervised detection of abnormal physiological signals. IEEE Journal of Biomedical and Health Informatics. ISSN: 2168-2194

<https://doi.org/10.1109/JBHI.2026.3715925>

© 2026 The Authors. Except as otherwise noted, this author-accepted version of a journal article published in IEEE Journal of Biomedical and Health Informatics is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Memory-Augmented Autoencoding with Self-Supervised Learning for Unsupervised Detection of Abnormal Physiological Signals

Zexin Wang, Lin Shi, Shoukun Xu, Menghui Zhou, Jianjun Chen, Po Yang, *Senior Member, IEEE* and Jun Qi, *Member, IEEE*

Abstract—Detecting abnormal events in physiological signals such as EEG and ECG is critical for early diagnosis of neurological and cardiovascular disorders. However, existing unsupervised anomaly detection methods often suffer from limited representation capacity and weak generalization across diverse signal domains. To address these challenges, we propose MAGE, a novel unsupervised anomaly detection framework that integrates multi-head memory gating, self-supervised learning, and adversarial training within a unified convolutional autoencoder architecture. The proposed memory-augmented gating mechanism selectively preserves and adaptively integrates salient features, improving discriminability and robustness over prior memory-based approaches. To further enhance representation learning, a self-supervised auxiliary task based on multiple signal transformations is introduced to encourage structure-aware feature extraction. In addition, transformation-aware adversarial perturbations are incorporated during training to enhance robustness against distribution shifts. Extensive experiments on multiple benchmark EEG and ECG datasets demonstrate that MAGE consistently outperforms state-of-the-art baselines in standard within-dataset evaluation settings, achieving over 98% detection accuracy and superior F1-scores. These results highlight the effectiveness and clinical potential of MAGE for early anomaly detection and continuous health monitoring under realistic deployment conditions. Code available at <https://github.com/wzxmodel/MAGE>.

Index Terms—Unsupervised anomaly detection, EEG, ECG, Self-supervised learning, Memory network

I. INTRODUCTION

ABNORMAL event detection in physiological signals, such as electroencephalograms (EEGs) and electrocardiograms (ECGs), is critical for early diagnosis and continuous patient monitoring. Given the high-dimensional, long-term, and highly variable nature of these signals in clinical settings,

manual inspection is inherently subjective and unscalable, driving the urgent need for automated anomaly detection.

While supervised learning achieves promising results, its reliance on large-scale, accurately annotated datasets limits its clinical viability, especially since physiological anomalies are rare and heterogeneous. Consequently, unsupervised anomaly detection has emerged as a highly practical alternative, identifying deviations by exclusively learning normal patterns from unlabeled data [1], [2]. This paradigm is particularly suited for long-term monitoring scenarios where annotations are scarce or unavailable.

Recent advances in deep learning, especially autoencoder-based models, have shown strong potential for unsupervised anomaly detection by modeling normal signal manifolds and using reconstruction error as an anomaly indicator [3], [4]. However, conventional autoencoders often struggle to capture both global physiological dynamics and subtle local variations, which are essential for detecting heterogeneous and transient abnormalities. As a result, their detection sensitivity and robustness under realistic clinical conditions remain limited.

To address these challenges, we propose MAGE, a novel unsupervised anomaly detection framework tailored for physiological signal analysis under a normality learning (one-class) setting, as illustrated in Fig. 1. Built upon a convolutional autoencoder backbone, MAGE introduces a Multi-Head Memory Gating (MHMG) module that adaptively integrates global and local feature representations through learnable gating mechanisms. This design enables selective memory retrieval and dynamic feature fusion, enhancing representation expressiveness while preserving computational efficiency. Gated residual connections further stabilize training and improve sensitivity to abnormal patterns.

In addition to architectural enhancements, MAGE incorporates a self-supervised auxiliary task that applies multiple signal transformations and trains the encoder to recognize them. This strategy encourages structure-aware and transformation-invariant feature learning without requiring manual labels. To further improve robustness under distributional shifts and signal perturbations commonly encountered in clinical data, a transformation-aware adversarial training scheme is introduced during optimization.

Extensive evaluations on multiple public EEG and ECG benchmarks demonstrate that MAGE consistently outperforms the latest state-of-the-art baselines, achieving superior detection accuracy and precision-recall performance. These results

Manuscript received 50 August 2025; revised 50 January 2025 and 50 April 2025; accepted 50 May 2025. Date of publication 50 May 2025; date of current version 50 August 2025. This work was supported in part by National Natural Science Foundation of China (62301452), in part by Suzhou Science and Technology Development Plan (SYW2025056), in part by Jiangsu Province Digital Twin Technology Engineering Research Center for Key Equipment of Petrochemical Process (Su Fa Gai Gao Ji Fa [2019] No.1125), and in part by Changzhou Key Laboratory of Urban Big Data Analysis and Application Technology (CM20193007). (Corresponding author: Jun Qi.)

Zexin Wang, Lin Shi and Shoukun Xu are with the Aliyun School of Big Data, Changzhou University, China. Zexin Wang and Lin Shi contribute equally.

Menghui Zhou and Po Yang are with the School of Computer Science, University of Sheffield, UK.

Jianjun Chen and Jun Qi are with the Department of Computing, Xi'an JiaoTong-Liverpool University, China (e-mail: jun.qi@xjtlu.edu.cn).

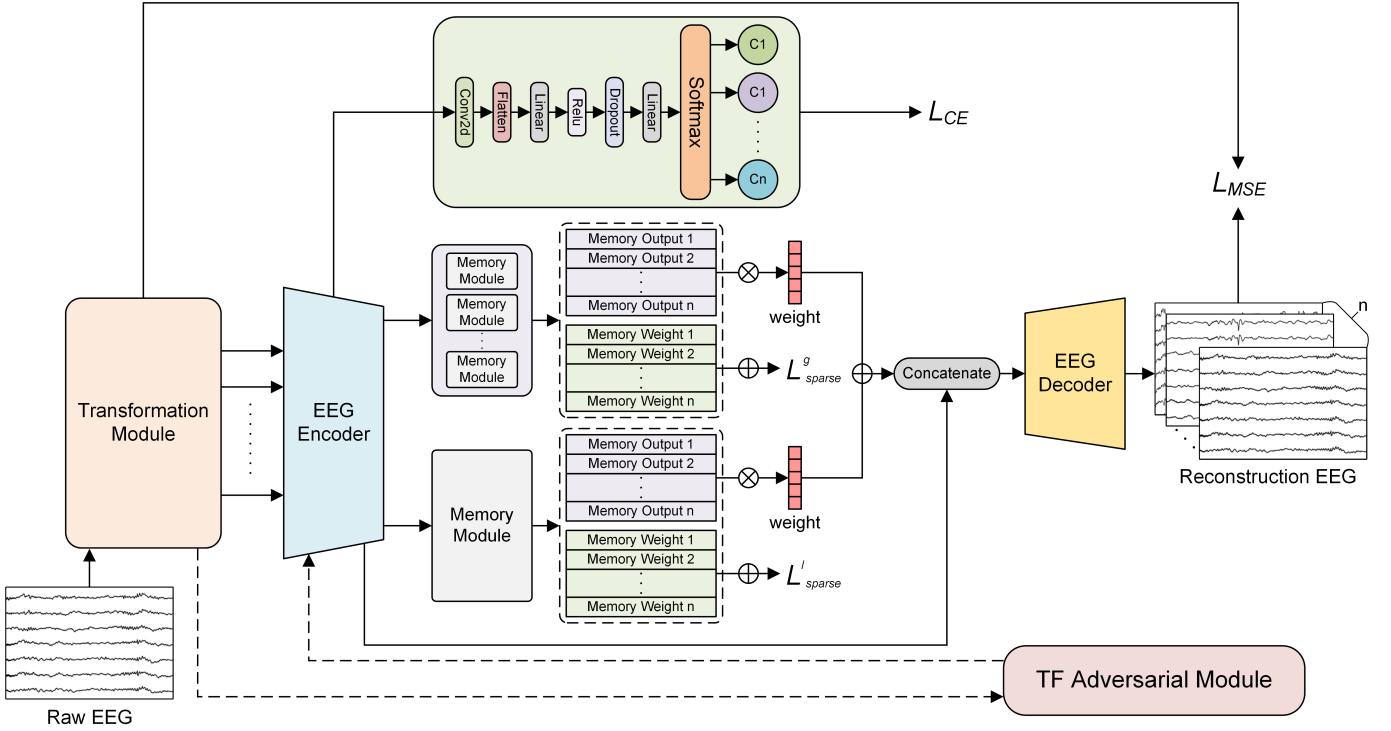


Fig. 1. Overall framework of the proposed MAGE. Built on a convolutional autoencoder backbone, MAGE integrates three key components: an Auxiliary Task Module for self-supervised feature regularization via transformation recognition, a Multi-Head Memory Gating module that adaptively retrieves and fuses representative normal patterns, and a Time-Frequency Adversarial Module that improves robustness through perturbations in both temporal and spectral domains. The model learns normal physiological patterns for unsupervised anomaly detection.

highlight the clinical viability and exceptional robustness of MAGE for early abnormality screening and continuous physiological monitoring.

The main contributions of this work are summarized as follows:

- We propose MAGE, an unsupervised anomaly detection framework that integrates a multi-head memory gating mechanism to enhance representation capacity and anomaly sensitivity in physiological signals.
- A self-supervised auxiliary learning strategy based on signal transformations is introduced to improve generalization without requiring anomaly labels.
- A transformation-aware adversarial training scheme is incorporated to enhance robustness against realistic signal perturbations and distributional shifts.
- Extensive experiments on multiple EEG and ECG benchmarks demonstrate the effectiveness and clinical applicability of the proposed method.

The remainder of this paper is organized as follows. Section II reviews related work. Section III presents the proposed method. Section IV describes the experimental setup. Section V reports and discusses the results, and Section VI concludes the paper.

II. RELATED WORK

In recent years, deep learning has developed rapidly [5], [6], showing substantial transformative potential across clinical medicine, particularly in tasks that demand high-precision

pattern recognition [7], [8]. Anomaly detection in physiological signals such as EEG and ECG, which serves as a cornerstone for early diagnosis of neurological and cardiovascular disorders, primarily relied on statistical analysis, spectral features, and hand-crafted descriptors in early studies [9]–[11]. Although these approaches provide interpretability, they struggle to model the nonlinear, non-stationary characteristics of physiological signals and are sensitive to noise and inter-subject variability.

With the development of deep learning, supervised models have achieved notable success in physiological signal analysis. Convolutional and recurrent neural networks are widely used for automatic feature extraction and temporal modeling [12]–[14]. Representative works include attention-based CNNs for seizure detection [15], hybrid convolution–recurrent architectures such as ChronoNet for EEG abnormality analysis [16], and RNN-based models for ECG arrhythmia classification [17]. Despite strong performance, these methods depend heavily on large-scale labeled datasets, which are costly and often unreliable in clinical practice.

To reduce reliance on annotations, unsupervised anomaly detection has gained increasing attention. Autoencoder-based methods are commonly adopted, using reconstruction error as an anomaly indicator. Variational and adversarial autoencoders have been proposed to enhance representation capacity and robustness [18], [19], with recent cutting-edge paradigms introducing deep LSTM-attention VAEs (LAVAE-Net) [?], sparse temporal autoencoder (STAE) [20], and dynamic graph-deviation networks (EGDN-KL) [?] for advanced EEG and

ECG analysis. However, these models often fail to jointly capture global physiological dynamics and subtle local variations efficiently without suffering from over-smoothing or computational bottlenecks, limiting their ability to detect heterogeneous or transient abnormalities.

Memory-augmented neural networks enhance anomaly detection by explicitly storing representative normal patterns [21], which improves reconstruction fidelity and anomaly separability in physiological signals [22], [23]. However, most existing approaches rely on static or single-head memory designs. These lack adaptability to diverse and evolving signal characteristics, thereby limiting their real-world generalization.

Self-supervised learning (SSL) effectively extracts robust representations without manual labels via auxiliary pre-text tasks [24]. In biomedical applications, SSL mitigates annotation scarcity and boosts anomaly detection performance through methods like Transformer-based epilepsy detection [25], auxiliary denoising autoencoders for ECG [26], and multi-transformation time-series frameworks [27]. These studies prove that leveraging intrinsic signal structures significantly enhances representation quality.

Adversarial learning has also been explored to improve model robustness against noise and distributional shifts, demonstrating enhanced resilience in EEG classification and anomaly detection [28], [29]. Nevertheless, these adversarial strategies are typically applied in isolation and are rarely integrated with memory mechanisms or SSL objectives.

In summary, existing studies often investigate memory modeling, self-supervised learning, or adversarial training independently. Few works integrate these complementary strategies into a unified framework for physiological anomaly detection. To bridge this gap, we propose MAGE, an unsupervised framework that jointly incorporates multi-head memory gating, transformation-based self-supervised learning, and adversarial training, enabling more expressive representation learning and improved robustness across heterogeneous physiological modalities.

In summary, existing studies predominantly investigate memory modeling, SSL, and adversarial training independently. To bridge this gap, we propose MAGE, a unified unsupervised framework that jointly integrates multi-head memory gating, transformation-based SSL, and adversarial training. This synergy enables more expressive representation learning and superior robustness across heterogeneous physiological modalities.

III. METHODOLOGY

This section describes the architecture and training strategy of the proposed MAGE model for unsupervised anomaly detection in physiological signals. As illustrated in Fig. 1, MAGE integrates a convolutional autoencoder backbone with a Multi-Head Memory Gating module, a self-supervised auxiliary task, and an adversarial training mechanism. These components operate jointly to capture both global structures and subtle abnormalities in complex biomedical time-series data such as EEG and ECG.

A. Overview

MAGE is built upon a convolutional autoencoder (CAE) for normality modeling, where anomalies are detected based on reconstruction error. The CAE adopts an encoder-decoder architecture that maps an input signal x to a latent representation z , and reconstructs it as:

$$\begin{aligned} z &= f_{enc}(x; \omega_{enc}), \\ \hat{x} &= f_{dec}(z; \omega_{dec}), \end{aligned} \quad (1)$$

where ω_{enc} and ω_{dec} denote the parameters of the encoder and decoder, respectively.

The reconstruction loss is measured using the mean squared error:

$$\mathcal{L}_{MSE} = \|x - \hat{x}\|_2^2. \quad (2)$$

Low reconstruction error indicates that the input conforms to the learned normal signal manifold, whereas anomalous inputs typically yield higher reconstruction errors. During inference, samples exceeding a predefined reconstruction threshold are identified as abnormal.

B. Auxiliary Task Module

The Auxiliary Task Module implements a self-supervised learning strategy to enhance representation quality and generalization under the normality-learning setting, where only normal samples are available for training. Since anomalous patterns are rare, diverse, and usually unseen during training, relying solely on reconstruction may lead to overfitting to superficial signal characteristics. To mitigate this issue, MAGE introduces a transformation recognition task as an auxiliary objective, encouraging the encoder to learn structured and discriminative latent representations without requiring manual labels.

Specifically, a predefined set of signal transformations is applied to each input signal, and the encoder is trained to identify the applied transformation. These transformations act as pseudo-labels and expose the model to diverse signal variations, promoting robustness and invariance. We define seven transformations $\{\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3, \mathcal{T}_4, \mathcal{T}_5, \mathcal{T}_6\}$, each targeting a different signal characteristic:

- 1) *Raw* ($\mathcal{T}_0$): Original signal without modification, serving as a baseline.
- 2) *Noise* ($\mathcal{T}_1$): Adds Gaussian noise to simulate sensor disturbances and improve noise robustness.
- 3) *Reverse* ($\mathcal{T}_2$): Reverses the temporal order of the signal to disrupt directional dependencies.
- 4) *Permute* ($\mathcal{T}_3$): Divides the signal into segments and randomly shuffles them to test temporal consistency.
- 5) *Scale* ($\mathcal{T}_4$): Multiplies the signal by a random scalar $s \in \{0.3, 0.8, 1.3, 1.8, 2.3\}$, promoting scale invariance.
- 6) *Dropout* ($\mathcal{T}_5$): Randomly select a time point and set its value to 0.
- 7) *Masking* ($\mathcal{T}_6$): Replaces a continuous time segment with a specific value.

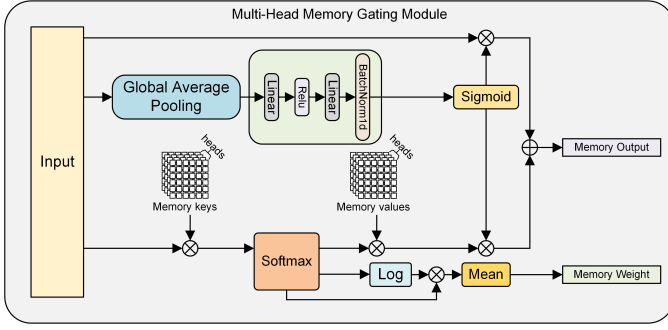


Fig. 2. Architecture of the Multi-Head Memory Gating Module. It introduces multiple memory heads and a learnable gating function to adaptively fuse memory-enhanced features with the original input.

Each transformed signal is encoded and classified according to its transformation type using a softmax classifier. The auxiliary objective is defined by the cross-entropy loss:

$$\mathcal{L}_{CE} = - \sum_{i=1}^{|\mathcal{T}|} y_i \log(p_i), \quad (3)$$

where $\mathcal{T}$ denotes the number of transformations, y_i is the one-hot label, and p_i is the predicted probability.

By jointly optimizing the reconstruction loss and the auxiliary transformation classification loss, the encoder is encouraged to capture structure-aware and semantically meaningful features. This joint learning strategy substantially improves sensitivity to subtle and heterogeneous anomalies in complex physiological time-series data.

C. Multi-Head Memory Gating Module

MHMG module enhances representation learning by introducing a structured memory mechanism into the convolutional autoencoder. Conventional autoencoders often overfit to normal training data and exhibit limited generalization to unseen anomalies. MHMG alleviates this issue by enabling selective retrieval of representative normal patterns from historical feature space and adaptive fusion with current features, as illustrated in Fig. 2.

Given an input feature tensor $X \in \mathbb{R}^{B \times C \times H \times W}$, the channel dimension is evenly divided into H_m memory heads, each operating on a subspace of dimension $d_h = C/H_m$. The reshaped input is expressed as:

$$\begin{aligned} X' &= \text{reshape}(X; (T_x, H_m, d_h)), \\ T_x &= B \times H \times W. \end{aligned} \quad (4)$$

Each memory head maintains learnable memory keys and values, $M_k, M_v \in \mathbb{R}^{H_m \times d_h \times C}$. For the h -th head, similarity scores between the input features and memory keys are computed using scaled dot-product attention:

$$D_h = \frac{X'_h \cdot M_k^h}{\tau}, \quad (5)$$

where τ is a learnable temperature parameter controlling the sharpness of attention distribution. The memory-enhanced feature is obtained via a softmax-weighted sum:

$$A_h = \text{softmax}(D_h) \cdot M_v^h, \quad (6)$$

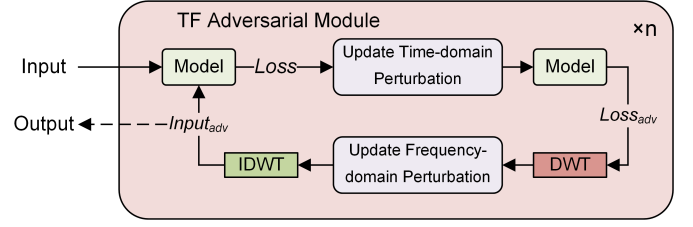


Fig. 3. Architecture of the Time-Frequency Adversarial Module. It introduces perturbations in both time and frequency domains to generate adversarial samples while preserving the intrinsic characteristics of the original signals.

resulting in $A_h \in \mathbb{R}^{B \times H \times W \times d_h}$. The outputs from all heads are concatenated to reconstruct the full feature space:

$$A = \text{concat}(A_1, A_2, \dots, A_{H_m}). \quad (7)$$

To adaptively combine the memory-enhanced representation A , with the original features X , a learnable gating mechanism is applied:

$$\begin{aligned} G &= \text{AvgPool}(X) \in \mathbb{R}^{B \times C}, \\ G' &= \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot G)), \\ \tilde{X} &= G' \odot A + (1 - G') \odot X, \end{aligned} \quad (8)$$

where $W_1 \in \mathbb{R}^{C \times C/3}$, $W_2 \in \mathbb{R}^{C \times C/3}$ are learnable projection matrices, and σ denotes the sigmoid activation. G' acts as a channel-wise gate to control the fusion, and $\tilde{X}$ is the final fused representation.

To encourage diverse memory access and prevent dominance of a few memory slots, an attention entropy regularization term is introduced:

$$H_{att} = - \sum_{i=1}^{N_m} A_{h,i} \cdot \log A_{h,i}, \quad (9)$$

where $A_{h,i}$ is the attention weight for the i -th memory slot in head h . The total regularization loss is:

$$\mathcal{L}_{\text{spar}} = \lambda \cdot H_{att}, \quad (10)$$

with λ as a balancing hyperparameter.

Overall, MHMG enables specialization across multiple feature subspaces while adaptively controlling information fusion. This design improves representation expressiveness, mitigates irrelevant feature interference, and enhances generalization in physiological anomaly detection.

D. TF Adversarial Module

Deep neural networks for physiological anomaly detection are vulnerable to adversarial perturbations, which can significantly degrade performance under real-world signal variability. To enhance robustness, MAGE incorporates a Time-Frequency Adversarial module that generates adversarial samples by applying constrained perturbations in both domains [30], as shown in Fig. 3. The perturbations are optimized to maximize the model's loss while preserving the structural and perceptual characteristics of the original signals.

Algorithm 1 TF-based Adversarial Attack Algorithm for Physiological Signal Perturbation

Input: Input data x , Number of attack iterations N , Model parameters ω , perturbation step size η , Batch size B , Number of data transformation types T , Number of sensor channels C , Sampling point N_t , Loss function $\mathcal{L}(\cdot)$, Forward propagation of the model $f(\cdot)$.

Output: Adversarial example x_{adv}

```

1: Initialize  $x_{\text{adv}} \leftarrow x$ 
2: for  $t \leftarrow 1$  to  $N$  do
3:   /* TIME DOMAIN ATTACK */
4:    $L_t \leftarrow \mathcal{L}(f(x_{\text{adv}}; \omega))$ 
5:    $g_t \leftarrow \text{sign}(\nabla_{x_{\text{adv}}} L_t)$ 
6:    $x_{\text{adv}} \leftarrow x_{\text{adv}} + \eta \cdot g_t$ 
7:    $x_{\text{adv}} \leftarrow \text{clip}(x_{\text{adv}}, 0, 1)$ 
8:   /* FREQUENCY DOMAIN ATTACK */
9:    $L_f \leftarrow \mathcal{L}(f(x_{\text{adv}}; \omega))$ 
10:   $g_f \leftarrow \text{sign}(\nabla_{x_{\text{adv}}} L_f)$ 
11:   $x_{\text{adv}} \leftarrow \text{reshape}(x_{\text{adv}}, B \cdot T \cdot 1 \cdot C, N_t)$ 
12:   $\text{coeffs} \leftarrow \text{DWT}(x_{\text{adv}})$ 
13:  for  $i \leftarrow 1$  to 4 do
14:     $\text{coeffs}_i \leftarrow \text{coeffs}_i + \eta \cdot g_f$ 
15:  end for
16:   $x_{\text{adv}} \leftarrow \text{IDWT}(\text{coeffs})$ 
17:   $x_{\text{adv}} \leftarrow \text{clip}(x_{\text{adv}}, 0, 1)$ 
18: end for
19: return  $x_{\text{adv}}$ 

```

1) *Time-Domain Perturbation:* In the time domain, adversarial perturbations are computed based on the gradient of the total loss with respect to the input signal. Given an original sample x , the adversarial version x_{adv} is updated iteratively using the following rule:

$$x_{\text{adv}}^{t+1} = x_{\text{adv}}^t + \eta \cdot \text{sign}(\nabla_x \mathcal{L}_{\text{total}}(x_{\text{adv}}^t)), \quad (11)$$

where η is the step size, $\text{sign}(\cdot)$ denotes the element-wise sign function, and $\mathcal{L}_{\text{total}}$ is the overall loss function of the MAGE model. This procedure ensures that the generated adversarial perturbations are aligned with the most sensitive directions of the model, thereby exposing its vulnerability while guiding robustness-oriented optimization.

2) *Frequency-Domain Perturbation:* To further enhance the effectiveness and diversity of adversarial perturbations, the attack is extended to the frequency domain using the Discrete Wavelet Transform (DWT). Specifically, the adversarial input x_{adv}^t is decomposed into four frequency sub-bands:

$$C = \text{DWT}(x_{\text{adv}}^t) = \{C_{ll}, C_{lh}, C_{hl}, C_{hh}\}, \quad (12)$$

where C_{ll} represents the low-frequency component, while C_{lh} , C_{hl} , and C_{hh} correspond to high-frequency components in different orientations.

Perturbations are computed based on the gradient of the total loss $\mathcal{L}_{\text{total}}$ with respect to the wavelet coefficients and are applied as follows:

$$\begin{aligned} C'_i &= C_i + \alpha \cdot \mathbb{E}_t[\text{sign}(\nabla_{x_{\text{adv}}} \mathcal{L}_{\text{total}}(x_{\text{adv}}))], \\ x_{\text{adv}}^{t+1} &= \text{IDWT}(C'), \end{aligned} \quad (13)$$

Algorithm 2 The procedure of the MAGE

Input: Input data $x \in \mathbb{R}^{B \times C \times N_t}$, Epoch E , Batch size B , Number of data transformation types $\mathcal{T} = \{\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3, \mathcal{T}_4, \mathcal{T}_5, \mathcal{T}_6\}$, Number of sensor channels C , Sampling point N_t , Cross-entropy loss function $\mathcal{L}_{\text{CE}}(\cdot)$, The parameters of the submodule ω_i , Forward propagation of the submodule $f_i(\cdot)$, hyperparameters λ_1, λ_2 .

Output: $L_{\text{mse}}, L_{\text{spar}}, L_{\text{ce}}, L_{\text{total}}$

```

1: for epoch  $\leftarrow 1$  to  $E$  do
2:    $X \leftarrow \text{Tran}(x), X \in \mathbb{R}^{B \times T \times 1 \times C \times N_t}$ 
3:    $Z \leftarrow f_{\text{encoder}}(X; \omega_{\text{enc}})$ 
4:    $G \leftarrow f_{\text{class}}(Z; \omega_{\text{class}})$ 
5:    $L_{\text{ce}} \leftarrow 0$ 
6:   for  $i \leftarrow 1$  to  $|\mathcal{T}|$  do
7:      $L_{\text{ce}} \leftarrow L_{\text{ce}} + \mathcal{L}_{\text{CE}}(G_i, i)$ 
8:   end for
9:    $(M_g, A_g) \leftarrow f_{\text{gm}}(X; \omega_{\text{gm}})$ 
10:   $(M_l, A_l) \leftarrow f_{\text{lm}}(X; \omega_{\text{lm}})$ 
11:   $\tilde{Z} \leftarrow \text{Fusion}(M_g, M_l; \omega_f)$ 
12:   $\hat{Z} \leftarrow \text{Concatenate}(Z, \tilde{Z})$ 
13:   $\hat{X} \leftarrow f_{\text{decoder}}(\hat{Z}; \omega_{\text{dec}}), \hat{X} \in \mathbb{R}^{B \times T \times 1 \times C \times N_t}$ 
14:   $L_{\text{mse}} \leftarrow \|X - \hat{X}\|^2$ 
15:   $L_{\text{spar}} \leftarrow \sum A_g + \sum A_l$ 
16:   $L_{\text{total}} \leftarrow L_{\text{mse}} + \lambda_1 \cdot L_{\text{spar}} + \lambda_2 \cdot L_{\text{ce}}$ 
17: end for
18: return  $L_{\text{mse}}, L_{\text{spar}}, L_{\text{ce}}, L_{\text{total}}$ 

```

where α is a scaling factor controlling the perturbation magnitude, and $\mathbb{E}_t[\cdot]$ denotes the expectation over time frames. The modified coefficients C'_i are subsequently transformed back to the time domain via inverse DWT (IDWT) to obtain the updated adversarial signal.

3) *Time Frequency-Domain Perturbation:* The proposed time-frequency (TF) adversarial perturbation strategy enhances anomaly detection in physiological signals by jointly introducing perturbations in both domains. This design increases the effectiveness of adversarial samples while preserving perceptual consistency. Unlike conventional attacks that optimize only classification loss, the TF-based strategy optimizes the overall objective, including auxiliary losses, thereby improving perturbation diversity and precision. The procedure is summarized in Algorithm 1.

By selectively adjusting frequency components through gradient-based optimization, this strategy preserves essential physiological structures while introducing imperceptible distortions. Consequently, TF-based perturbations provide a realistic assessment of model vulnerability and facilitate robustness-oriented training.

During inference, anomalies are identified by thresholding reconstruction errors. The memory-guided adaptive feature fusion further enables dynamic adjustment to unseen patterns, improving generalization. The complete inference procedure is described in Algorithm 2. Overall, the integration of TF-based adversarial training and memory-guided inference enhances robustness and adaptability while maintaining computational efficiency.

TABLE I
SUMMARY OF EEG AND ECG BENCHMARK DATASETS USED IN THE
ANOMALY DETECTION EXPERIMENTS

Dataset	TUSZ	CHB-MIT	AUBMC	INCART	PTB
Sessions / Records	1643	198	35	75	549
Subjects	675	22	6	32	290
Sampling Rate (Hz)	256	256	500	257	1000
Channels / Dimensions	19	23	19	12	15
Train Samples (Normal Only)	15780	14278	4700	10048	9453
Anomalies (%)	4.65	0.27	1.91	14.43	16.38

IV. DATASET AND EXPERIMENTAL SETUP

To rigorously assess the effectiveness of the proposed MAGE model in physiological signal anomaly detection, we conduct experiments on multiple benchmark EEG and ECG datasets. These datasets encompass diverse pathological conditions, enabling a comprehensive evaluation of the model's generalization and robustness.

A. EEG datasets

We evaluate the proposed anomaly detection framework using three publicly accessible EEG datasets, namely TUSZ, CHB-MIT, and AUBMC. These datasets collectively encompass large-scale adult recordings, pediatric epilepsy cases, and small-cohort datasets with clinical annotations. A unified statistical summary of these datasets, encompassing subject counts, recording characteristics, and anomaly ratios, is presented in Table I.

The Temple University Hospital EEG Corpus (TUSZ) is a large-scale clinical dataset containing recordings from 675 subjects, with more than 3,500 annotated seizure events spanning multiple seizure categories [31], [32]. The total recording duration exceeds 5 million seconds, while seizure activity constitutes only a small fraction of the data. Following standard practice, segments labeled as background activity are treated as normal signals, whereas all seizure-related segments are regarded as anomalous events. The richness of seizure types and the extreme class imbalance make TUSZ a challenging benchmark for unsupervised anomaly detection.

The CHB-MIT Scalp EEG Database consists of long-term EEG recordings from 22 pediatric patients with intractable epilepsy, sampled at 256 Hz using up to 23 scalp electrodes following the international 10–20 system. A total of 198 seizures are annotated across 129 seizure-containing sessions [33], [34]. In our experiments, non-seizure EEG segments are used exclusively to model normal brain activity, while seizure segments are reserved for evaluation as anomalies. Despite minor recording gaps, CHB-MIT provides a clinically realistic and widely adopted benchmark for pediatric seizure detection.

The AUBMC dataset was collected from six patients with focal epilepsy undergoing long-term video-EEG monitoring at the American University of Beirut Medical Center. EEG signals are recorded at 500 Hz using 19 scalp electrodes, with detailed annotations describing seizure onset zones, types, and durations [35]. Each recording contains interictal segments and one or more ictal events, resulting in over one hour of cumulative seizure activity. Consistent with the other EEG

datasets, non-seizure segments are considered normal data, while all seizure-related segments are treated as anomalous samples in our experiments.

B. ECG datasets

We further evaluate the proposed framework on two widely used ECG benchmarks, INCART and PTB, which represent long-term Holter monitoring data and high-resolution clinical diagnostic recordings, respectively. Key dataset statistics and sample distributions are summarized in Table I.

The St Petersburg INCART 12-lead Arrhythmia Database contains 75 annotated 30-minute ECG recordings obtained from 32 patients undergoing clinical evaluation for coronary artery disease. Signals are recorded at 257 Hz using 12 standard leads, with over 175,000 beat-level annotations covering a broad range of cardiac abnormalities, including myocardial infarction, atrial fibrillation, supraventricular ectopy, conduction disorders, and bundle branch blocks [36]. In our experiments, beats labeled as normal sinus rhythm are used exclusively for training, while all non-normal beats are treated as anomalous events during evaluation, enabling clinically meaningful anomaly detection from normal cardiac activity.

The PTB Diagnostic ECG Database consists of 549 high-resolution ECG recordings from 290 subjects, acquired at 1000 Hz using 15 synchronized leads (12 standard leads and 3 Frank leads). Each subject contributes one to five recordings, accompanied by detailed clinical metadata spanning multiple diagnostic categories, with myocardial infarction being the most prevalent [36], [37]. In our anomaly detection setting, recordings from healthy subjects are used as normal data for training, whereas recordings associated with any pathological diagnosis are regarded as anomalies during evaluation.

C. Experimental Setup

To ensure a consistent, clinically meaningful, and reproducible evaluation, we adopt a normality learning (one-class anomaly detection) paradigm across all EEG and ECG experiments. Specifically, the model is trained exclusively on normal physiological signals, while both normal and abnormal segments are included during testing to assess anomaly detection performance under realistic clinical conditions.

For EEG datasets, normal data consist of non-seizure segments, including interictal intervals from epileptic patients as well as recordings from healthy subjects. Abnormal data correspond to ictal segments characterized by rhythmic discharges, abrupt amplitude changes, and abnormal synchronization patterns, in accordance with standard clinical definitions of epileptic seizures. For ECG datasets, normal signals are derived from subjects without diagnosed cardiac abnormalities and typically exhibit regular sinus rhythms, whereas abnormal segments include myocardial infarction, arrhythmias, conduction disorders, and ischemic events, which often manifest as waveform irregularities or ST-segment deviations.

All recordings are segmented into non-overlapping 1-second windows for normal data. This window length was empirically selected as a practical trade-off between preserving sufficient

TABLE II
PERFORMANCE COMPARISON OF MAGE WITH STATE-OF-THE-ART METHODS ON EEG DATASETS

Method	TUSZ				CHB-MIT				AUBMC			
	Acc(%)	Pre(%)	Rec(%)	F1-score	Acc(%)	Pre(%)	Rec(%)	F1-score	Acc(%)	Pre(%)	Rec(%)	F1-score
C-LSTM [43]	64.99	75.86	72.00	0.7387	69.73	69.56	69.43	0.6949	70.48	71.94	80.80	0.7612
ChronoNet [44]	78.33	79.03	84.50	0.8167	81.65	82.19	80.81	0.8149	86.82	87.21	89.14	0.8817
ADSA [45]	82.61	81.79	89.50	0.8547	89.88	89.63	90.20	0.8991	88.94	89.06	91.12	0.9008
USAD [46]	73.50	75.17	80.08	0.7755	88.25	87.12	89.79	0.8843	82.73	89.98	77.25	0.8313
CAE-M [47]	66.51	73.81	64.18	0.6865	64.34	67.48	55.36	0.6082	64.51	71.30	59.51	0.6488
Anomaly Transformer [48]	93.29	96.08	92.00	0.9400	94.49	94.95	93.99	0.9447	98.11	98.45	98.10	0.9828
BeatGAN [49]	87.94	89.67	89.18	0.8942	88.84	89.49	88.03	0.8875	84.30	86.16	85.19	0.8567
TimesNet [50]	89.57	91.60	90.00	0.9079	86.31	86.29	86.34	0.8632	94.27	95.07	94.51	0.9479
LCT [51]	94.79	95.18	95.72	0.9545	95.79	96.30	95.23	0.9576	96.15	96.58	96.43	0.9651
GATs-LSTM [52]	93.36	95.05	93.22	0.9413	94.97	94.74	95.23	0.9498	92.22	93.18	92.66	0.9292
CwA-T [53]	84.79	85.59	88.22	0.8689	88.43	88.24	88.70	0.8847	83.74	85.44	84.96	0.8520
Fd-CAE [54]	92.23	94.56	91.68	0.9309	93.74	94.47	92.93	0.9369	88.55	93.44	85.19	0.8912
SAnoDDPM [55]	97.86	98.00	98.25	0.9813	97.68	97.37	98.01	0.9769	96.08	96.58	96.30	0.9644
SGAT-AE [56]	96.54	96.86	97.10	0.9698	98.11	97.87	98.37	0.9812	95.93	96.30	96.30	0.9630
DWT-CSP-CatBoost [57]	96.39	96.85	96.82	0.9684	97.84	97.86	97.83	0.9784	96.10	96.31	96.61	0.9646
STAE [20]	80.29	84.80	79.80	0.8223	82.74	84.79	79.79	0.8222	81.00	84.80	79.79	0.8222
EGDN-KL [58]	86.77	90.05	86.40	0.8819	88.14	88.37	88.37	0.8817	90.17	91.84	90.17	0.9100
LAVAE-Net [59]	90.29	91.09	92.00	0.9154	86.50	85.44	88.00	0.8670	92.65	92.75	93.99	0.9337
MAGE	98.29	98.60	98.40	0.9850	98.80	98.82	98.77	0.9880	98.64	98.99	98.54	0.9876
Improvement $\uparrow$	0.43	0.60	0.15	0.0037	0.69	0.95	0.40	0.0068	0.53	0.54	0.44	0.0048

temporal context to capture characteristic EEG and ECG patterns and maintaining computational efficiency, as shorter windows yield unstable representations while longer windows provide no consistent performance gains [38], [39]. To better capture the transient and sparse nature of pathological events, abnormal segments (seizure or cardiac events) are further processed using a sliding window strategy with 50% overlap during testing. The training set is constructed using only normal windows, while the test set combines normal and abnormal windows, reflecting a clinically realistic deployment scenario in which anomalies are rare, unseen during training, and only identified during inference. Importantly, anomaly labels are never used during training and are employed solely for evaluation.

We adopt a sample-level split strategy, which is suitable for unsupervised anomaly detection focused on learning shared normal physiological patterns rather than subject-specific features. Enforcing strict subject-wise separation in preliminary experiments substantially reduced normal data diversity and led to unstable training. This setting also aligns with realistic clinical monitoring scenarios, where models are trained on heterogeneous normal data and deployed across subjects [40]–[42].

Channel-wise Z-score normalization is performed on a per-recording basis, where the mean and standard deviation are computed independently for each channel within each EEG or ECG recording. This preprocessing preserves relative temporal-spatial patterns while ensuring consistent input scaling and stable model convergence across heterogeneous datasets.

To determine suitable model parameters, we evaluated multiple configurations. The memory matrix size was tested with values 32, 64, 128, 256, and the number of attention heads in the multi-head memory gating module was varied among 2, 4, 8, 16. The temperature parameter was assessed within 0.1, 0.3, 0.5, 0.7, 0.9. Optimal performance was achieved with a memory size of 64, four attention heads, and a temperature of

TABLE III
PERFORMANCE COMPARISON OF MAGE WITH STATE-OF-THE-ART METHODS ON ECG DATASETS

Method	INCART				PTB			
	Acc(%)	Pre(%)	Rec(%)	F1-score	Acc(%)	Pre(%)	Rec(%)	F1-score
C-LSTM [43]	78.45	78.26	81.39	0.7980	77.10	78.77	74.20	0.7642
ChronoNet [44]	87.55	86.62	90.09	0.8832	87.10	85.60	89.20	0.8737
ADSA [45]	88.32	87.15	91.09	0.8907	88.62	87.30	90.40	0.8882
USAD [46]	87.59	87.10	89.52	0.8829	82.28	80.93	84.45	0.8265
CAE-M [47]	70.18	71.88	70.57	0.7122	66.42	65.80	68.40	0.6708
Anomaly Transformer [48]	96.36	96.36	96.70	0.9653	97.80	97.05	98.60	0.9782
BeatGAN [49]	88.36	89.59	87.96	0.8877	88.92	89.34	88.40	0.8887
TimesNet [50]	90.02	90.87	89.96	0.9041	90.02	89.97	90.10	0.9003
LCT [51]	84.64	85.96	84.39	0.8517	81.58	81.88	81.10	0.8149
GATs-LSTM [52]	85.59	86.57	85.74	0.8615	83.68	83.62	83.75	0.8369
CwA-T [53]	76.50	77.79	77.04	0.7741	71.80	72.61	70.00	0.7128
Fd-CAE [54]	86.09	89.11	83.61	0.8627	86.42	88.77	83.40	0.8600
SAnoDDPM [55]	94.14	96.28	92.35	0.9427	95.12	95.28	94.95	0.9512
SGAT-AE [56]	93.18	95.45	91.30	0.9333	95.25	95.48	95.00	0.9524
DWT-CSP-CatBoost [57]	93.73	94.82	93.09	0.9394	96.58	97.02	96.10	0.9656
STAE [20]	81.95	84.80	79.78	0.8221	82.75	84.80	79.80	0.8223
EGDN-KL [58]	93.23	93.11	94.00	0.9355	96.13	95.78	96.50	0.9614
LAVAE-Net [59]	94.77	94.04	96.09	0.9305	96.55	95.77	97.40	0.9658
MAGE	98.91	99.13	98.78	0.9895	98.52	98.45	98.60	0.9853
Improvement $\uparrow$	2.55	2.77	2.08	0.0242	0.72	1.40	0.0	0.0071

0.7.

Hyperparameter tuning was conducted via a combination of grid search and manual refinement. Learning rates in the range of 1×10^{-6} to 1×10^{-4} are evaluated, with 1×10^{-5} providing the best validation performance. Batch sizes of 16, 32, and 64 are tested, and 32 was selected for its balance between efficiency and generalization. A dropout rate of 0.5 was adopted to mitigate overfitting. More specific parameter settings will be listed in the experimental results section.

All models were trained for 120 epochs to ensure convergence. The total loss function comprised three components: mean squared error (L_{mse}) for reconstruction accuracy, sparsity regularization (L_{spar}) to promote compact latent representations, and a cross-entropy term (L_{ce}) derived from the self-supervised auxiliary task to promote transformation-level discrimination. The overall loss function is defined as:

$$L_{total} = L_{mse} + \lambda_1 \cdot L_{spar} + \lambda_2 \cdot L_{ce}, \quad (14)$$

where λ_1 and λ_2 are hyperparameters that control the contributions of the regularization terms. In our experiments, λ_1

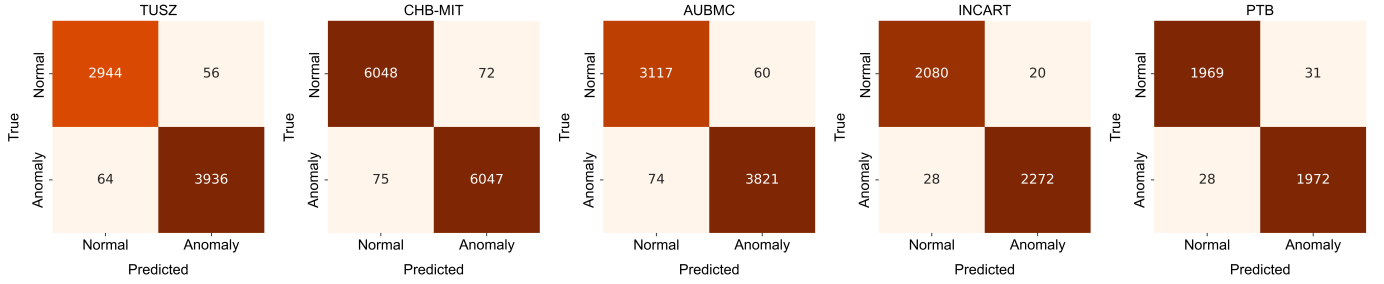


Fig. 4. Confusion matrices of MAGE on EEG and ECG benchmark test sets. Rows represent ground-truth labels and columns represent predictions. Balanced visualization is used for clarity, while evaluation follows the original data distributions.

was set to 1.0 and λ_2 to 0.0002. The proposed model was implemented in PyTorch and trained using mixed-precision on an NVIDIA RTX 3090 GPU to improve computational efficiency. Performance was evaluated using standard anomaly detection metrics, including F1-score and accuracy, to assess the model’s effectiveness in detecting abnormalities across diverse EEG and ECG datasets.

V. RESULTS AND DISCUSSION

The effectiveness of MAGE is evaluated through comprehensive comparisons with state-of-the-art unsupervised anomaly detection methods on five benchmark EEG and ECG datasets. All results are reported on held-out test sets, with training conducted using only the corresponding training data as described in Section IV-C. The baselines include statistical methods, autoencoder-based models, and recent memory-augmented and contrastive learning approaches for physiological signals.

As shown in Table II and Table III, MAGE consistently outperforms all baselines in accuracy, precision, recall, and F1-score across all datasets, achieving detection rates above 98% on both EEG and ECG benchmarks. These results demonstrate strong generalization across modalities, recording conditions, and clinical scenarios.

Further insights into the model’s detection behavior are provided by the confusion matrices shown in Fig. 4, which are computed solely on the test sets. These matrices summarize the distributions of true positives, true negatives, false positives, and false negatives, offering a detailed view of MAGE’s ability to discriminate between normal and abnormal segments. Although the matrices are displayed in a visually balanced format for clarity, this does not reflect the natural class distributions of the datasets. The original test set compositions are fully preserved, and no artificial balancing is applied during training or evaluation. As observed, MAGE maintains high sensitivity to pathological events while achieving low false positive rates, a desirable property for reducing false alarms in clinical monitoring applications.

The training dynamics and optimization stability of MAGE are further analyzed by examining the evolution of its loss components over 120 epochs, as shown in Fig. 5. The loss of sparsity regularization decreases rapidly during early training, indicating efficient adaptation of the memory module, while the reconstruction and auxiliary losses decline more gradually

as the model progressively refines its representations. The total loss exhibits a smooth and stable downward trend throughout training, confirming that the joint optimization of multiple objectives proceeds in a balanced manner without compromising convergence. This behavior demonstrates the stability of the proposed training strategy and its suitability for robust unsupervised anomaly detection.

Anomaly detection performance is further evaluated using receiver operating characteristic (ROC) and precision–recall (PR) curves across all datasets, as shown in Fig. 6.

The ROC curves consistently approach the top-left corner, with AUC-ROC values exceeding 0.99, indicating strong separability between normal and anomalous segments. Similarly, the PR curves maintain high precision and recall across decision thresholds, yielding large AUC-PR values even under class imbalance. This behavior is particularly important for clinical applications, where both missed anomalies and false alarms must be minimized, and demonstrates the strong generalization capability of MAGE.

The robustness of MAGE to hyperparameter variations is examined through a parameter sensitivity analysis on the TUSZ dataset, as shown in Fig. 7. Model performance remains stable across a broad range of values, with consistently high accuracy when the perturbation step size and the number of memory heads fall within reasonable bounds. This insensitivity to minor parameter changes is desirable for practical clinical deployment, where extensive hyperparameter tuning may be infeasible.

The impact of the decision threshold on detection performance is further analyzed using the TUSZ and INCART datasets. In MAGE, a test sample is classified as anomalous if its reconstruction error exceeds the $(100 \times q)$ -th percentile of the error distribution computed from the normal training data. As reported in Table IV, lower thresholds improve recall at the cost of increased false positives, whereas a higher threshold ($q = 0.99$) provides a more balanced trade-off between sensitivity and specificity, achieving the highest F1-score on both datasets. These results highlight the importance of threshold selection in real-world clinical scenarios, where the relative costs of false alarms and missed detections must be carefully considered.

Although MAGE is trained under a normality-learning paradigm, completely clean training data are rarely available in clinical practice due to annotation uncertainty, signal artifacts,

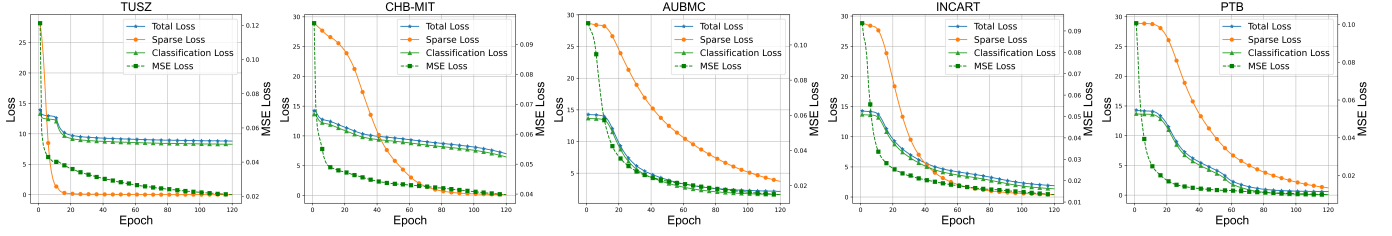


Fig. 5. Training loss curves of MAGE on representative EEG and ECG datasets. A stable and smooth convergence trend is observed across both modalities, indicating effective optimization and consistent training dynamics under the proposed multi-module framework.

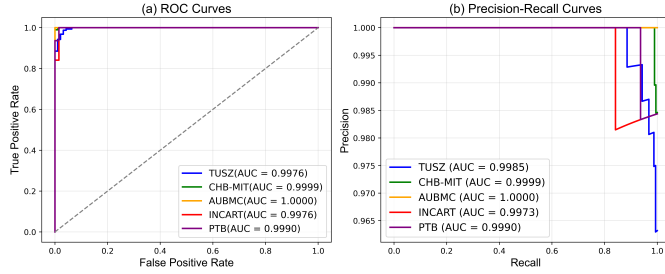


Fig. 6. Performance curves of MAGE on datasets. (a) AUC-ROC curves illustrating the true positive rate versus false positive rate. (b) Precision-Recall curves showing the trade-off between precision and recall. Each line represents a different dataset.

TABLE IV
PERFORMANCE OF MAGE UNDER DIFFERENT THRESHOLD VALUES q ON TUSZ AND INCART DATASETS

Threshold q	TUSZ			INCART		
	Acc(%)	Pre(%)	F1-score	Acc(%)	Pre(%)	F1-score
0.95	96.59 $\pm$ 0.15	97.39 $\pm$ 0.14	0.9700 $\pm$ 0.0013	96.87 $\pm$ 0.13	97.14 $\pm$ 0.19	0.9700 $\pm$ 0.0012
0.96	97.39 $\pm$ 0.10	98.04 $\pm$ 0.12	0.9771 $\pm$ 0.0008	97.94 $\pm$ 0.07	98.07 $\pm$ 0.05	0.9803 $\pm$ 0.0007
0.97	97.84 $\pm$ 0.07	98.30 $\pm$ 0.14	0.9811 $\pm$ 0.0006	98.44 $\pm$ 0.14	98.45 $\pm$ 0.22	0.9851 $\pm$ 0.0013
0.98	97.86 $\pm$ 0.06	98.37 $\pm$ 0.09	0.9812 $\pm$ 0.0005	98.12 $\pm$ 0.07	98.16 $\pm$ 0.12	0.9820 $\pm$ 0.0006
0.99	98.29$\pm$0.02	98.60$\pm$0.02	0.9850$\pm$0.0002	98.92$\pm$0.03	99.13$\pm$0.04	0.9896$\pm$0.0003

and retrospective data collection. To evaluate robustness under such realistic conditions, we progressively injected anomalous segments into the training set while keeping the training process fully unlabeled.

As shown in Fig. 8, detection performance on both EEG (TUSZ) and ECG (INCART) datasets degrades gradually as the proportion of anomalous samples increases. This behavior reflects the typical sensitivity of reconstruction-based anomaly detection methods to contamination in the normal data distribution, rather than indicating a strict requirement for anomaly-free training data. Importantly, anomaly labels are never used during training, and the contamination ratio is considered only for post hoc analysis.

The examined contamination levels are consistent with practical physiological monitoring scenarios, where EEG recordings typically contain sparse pathological events and ECG recordings may exhibit higher abnormal prevalence. The relatively stable performance under moderate contamination suggests that MAGE remains applicable when mild training data noise is unavoidable.

As shown in Table V, replacing the convolutional autoencoder backbone with a Transformer-based variant leads to consistent performance degradation and a substantial increase

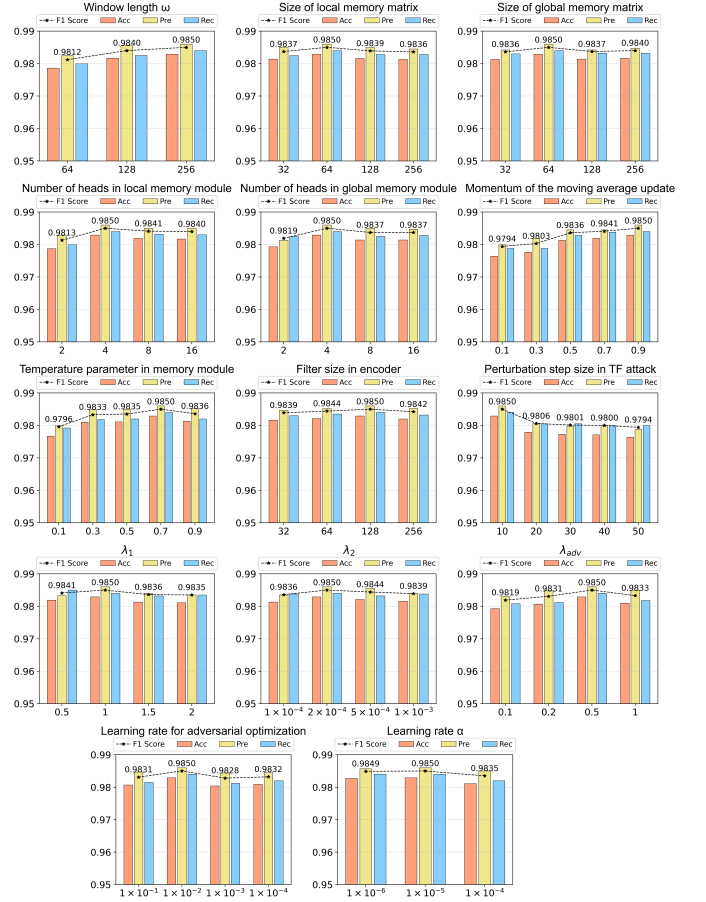


Fig. 7. Parameter sensitivity analysis of MAGE on the TUSZ dataset. Model performance remains stable across different values of key hyperparameters, demonstrating robustness to parameter variations.

in latency, indicating that the Transformer design is less suitable for efficient clinical deployment. Similarly, substituting the Multi-Head Memory Gating module with a single-head memory results in lower accuracy and F1-score, confirming the importance of multi-head memory for capturing heterogeneous physiological patterns.

We further compare different adversarial strategies. Replacing the proposed time-frequency adversarial training with FGSM [60], PGD [61], or C&W [62] attacks consistently degrades performance, while removing adversarial training entirely causes a statistically significant drop ($p < 0.05$). In addition, removing either the global or local memory component leads to highly significant performance degradation

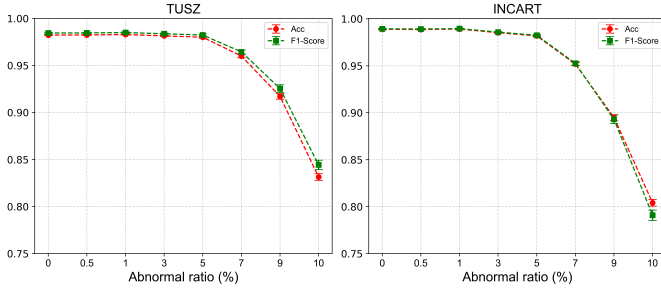


Fig. 8. Effect of anomalous sample contamination in training data on detection performance. Left: EEG dataset (TUSZ), Right: ECG dataset (INCART). Each subplot shows accuracy and F1-score with varying anomaly ratios in training data (0%, 0.5%, 1%, 3%, 5%, 7%, 9%, 10%).

TABLE V
PERFORMANCE COMPARISON OF MAGE AND ITS VARIANTS ON THE TUSZ AND INCART DATASETS

Variant	TUSZ			INCART		
	Acc(%)	Pre(%)	F1-score	Acc(%)	Pre(%)	F1-score
MAGE	98.29±0.02	98.60±0.02	0.9850±0.0002	98.92±0.03	99.13±0.04	0.9896±0.0003
MAGE-Transformer	97.65±0.26	97.96±0.34	0.9794±0.0023	97.90±0.10	98.16±0.04	0.9799±0.0009
MAGE-SHMem [63]	97.39±0.25	98.00±0.66	0.9771±0.0021	97.92±0.09	98.24±0.18	0.9801±0.0009
MAGE-FGSM	96.88±0.04	97.60±0.09	0.9726±0.0003	96.08±0.21	94.67±0.20	0.9631±0.0020
MAGE-PGD	97.45±0.04	98.22±0.17	0.9776±0.0003	97.28±0.05	95.85±0.03	0.9744±0.0005
MAGE-C&W	97.45±0.20	98.29±0.23	0.9775±0.0018	97.30±0.10	95.53±0.20	0.9747±0.0010
MAGE w/o Adv	96.26±0.16	97.64±0.12	0.9670±0.0014*	96.10±0.42	96.40±0.38	0.9626±0.0041*
MAGE w/o GlobalMem	95.46±0.26	95.69±0.39	0.9604±0.0024**	94.78±0.12	95.13±0.10	0.9500±0.0012**
MAGE w/o LocalMem	94.89±0.29	95.65±0.35	0.9552±0.0027**	94.67±0.28	95.02±0.20	0.9490±0.0027**
MAGE w/o AuxTask	91.12±0.88	93.45±0.59	0.9212±0.0080**	89.82±0.45	90.61±0.44	0.9022±0.0044**

* Statistically significant difference compared with the MAGE model ($p < 0.05$).

** Highly significant difference compared with the MAGE model ($p < 0.01$), based on the Wilcoxon signed-rank test on F1-score.

TABLE VI
COMPUTATIONAL COMPLEXITY AND INFERENCE LATENCY ON THE TUSZ AND INCART DATASETS

Variant	TUSZ		INCART	
	FLOPs (G)	Latency (ms)	FLOPs (G)	Latency (ms)
MAGE	6.99	43.59	5.15	96.45
MAGE-Transformer	6.67	182.34	4.32	412.53
MAGE-SHMem	7.03	46.29	4.74	97.11
Anomaly Transformer	10.02	100.98	11.84	99.47
SAnoDDPM	15.31	203.1	16.67	346.1
SGAT-AE	14.98	216.01	13.03	199.69
DWT-CSP-CatBoost	6.42	88.36	5.63	100.91

TABLE VII
EFFECT OF THE NUMBER OF SSL TRANSFORMATIONS ON MODEL PERFORMANCE ON THE TUSZ AND INCART DATASETS

Top-k	TUSZ				INCART			
	Acc(%)	Pre(%)	Rec(%)	F1-score	Acc(%)	Pre(%)	Rec(%)	F1-score
Top-1	93.06±0.18	94.71±0.15	93.06±0.41	0.9387±0.0018	93.50±0.20	94.05±0.18	93.48±0.23	0.9376±0.0020
Top-3	95.50±0.14	96.54±0.20	95.55±0.14	0.9604±0.0012	95.50±0.05	95.87±0.05	95.51±0.05	0.9569±0.0004
Top-5	97.12±0.03	97.76±0.05	97.18±0.07	0.9747±0.0003	97.50±0.06	97.66±0.13	97.55±0.03	0.9761±0.0006
All	98.29±0.02	98.60±0.02	98.40±0.01	0.9850±0.0002	98.91±0.03	99.13±0.04	98.78±0.02	0.9895±0.0003

($p < 0.01$), highlighting their complementary roles. Eliminating the self-supervised auxiliary task results in the largest performance decline, demonstrating its critical contribution to discriminative representation learning.

We select four representative baselines that consistently demonstrate strong detection performance, together with two architectural variants of the proposed framework for comparison. As shown in Table VI, MAGE attains comparable or lower FLOPs with substantially reduced inference latency compared

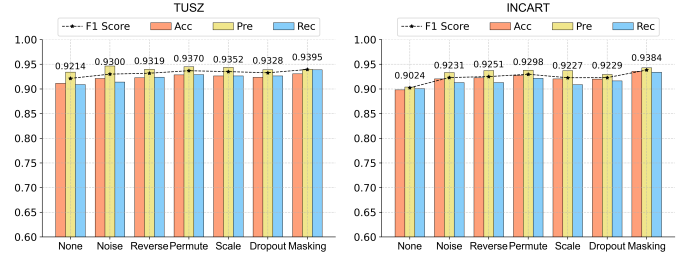


Fig. 9. Performance of different self-supervised transformations (used individually) on the main model over TUSZ and INCART datasets.

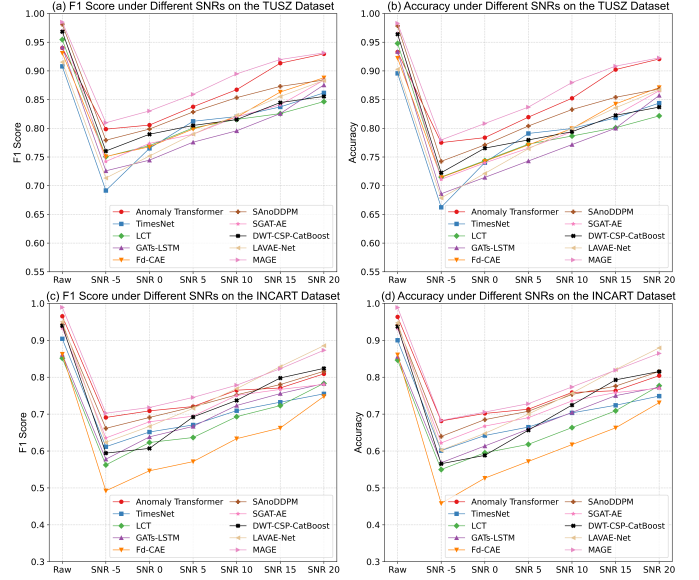


Fig. 10. Performance of different SNR on the TUSZ and INCART datasets.

to strong baselines. The higher latency of the Transformer-based variant, despite similar FLOPs, highlights the gap between theoretical complexity and practical efficiency. This demonstrates that the proposed convolutional backbone with memory gating offers a favorable balance between efficiency and deployability.

The effect of individual self-supervised transformations was evaluated by training the model with a single transformation at a time. As shown in Fig. 9, different transformations exhibit varying effectiveness, indicating that transformation selection plays a critical role in enhancing anomaly detection performance.

Building on these results, we further examined the impact of using multiple transformations. As reported in Table VII, performance improves consistently as more effective transformations are incorporated, with the best results achieved when all transformations are used. This demonstrates that combining diverse transformations provides stronger regularization for representation learning.

The noise robustness of MAGE was assessed by injecting Gaussian noise at SNR levels of -5 dB, 0 dB, 5 dB, 10 dB, 15 dB and 20 dB into raw EEG signals from the TUSZ dataset and raw ECG signals from the INCART dataset. The signal-

TABLE VIII
QUALITATIVE COMPARISON OF REPRESENTATIVE ANOMALY DETECTION METHODS AND THE PROPOSED MAGE FRAMEWORK IN TERMS OF SUPERVISION REQUIREMENTS, MODALITY COVERAGE, ROBUSTNESS, EFFICIENCY, INTERPRETABILITY, AND GENERALIZATION CAPABILITY.

Method	Label-Free	Anomaly	EEG	ECG	Noise	Contam.	Efficient	Explainable	Cross-Subject
GATs-LSTM	×	✓	✓	×	✓	×	×	✓	×
DWT-CSP-CatBoost	×	✓	✓	×	✓	×	✓	✓	✓
Anomaly Transformer	✓	✓	×	×	×	×	×	✓	×
SGAT-AE	✓	✓	×	×	✓	×	✓	✓	×
SAnoDDPM	✓	✓	✓	×	×	✓	✓	✓	×
LVAE-Net	✓	✓	×	✓	✓	×	✓	✓	×
MAGE (Ours)	✓	✓	✓	✓	✓	✓	✓	×	×

to-noise ratio (SNR) in decibels is defined as:

$$\text{SNR}_{\text{dB}} = 10 \cdot \log_{10} \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right), \quad (15)$$

$$P_{\text{noise}} = \frac{P_{\text{signal}}}{10^{\text{SNR}_{\text{dB}}/10}},$$

where P_{signal} and P_{noise} denote the power of the signal and the noise, respectively.

As shown in Fig. 10, performance decreases as noise intensity increases; however, MAGE consistently outperforms competing methods across all SNR levels. Even at -5 dB, MAGE achieves an F1-score of 0.8097 on TUSZ and 0.7026 on INCART, demonstrating strong robustness under severe noise contamination and suitability for real-world biomedical applications. In addition, as shown in Table VIII, MAGE combines label-free learning, multi-modal applicability, robustness, and computational efficiency within a unified framework. Although interpretability and cross-subject generalization require further investigation, MAGE offers a favorable balance between detection capability and practical deployability.

VI. CONCLUSION

In this study, we proposed MAGE, a memory-augmented anomaly detection framework for physiological time-series data that is tailored to realistic clinical settings. By integrating multi-head memory gating, self-supervised auxiliary learning, and time-frequency adversarial training within an autoencoding architecture, MAGE effectively captures stable normal physiological patterns while remaining robust to noise, data contamination, and distribution shifts commonly observed in clinical recordings.

Extensive evaluations on multiple EEG and ECG benchmarks demonstrate that MAGE achieves strong and consistent detection performance without requiring anomaly labels during training. The proposed framework exhibits notable robustness under contaminated training conditions, low signal-to-noise ratios, and inter-subject variability, making it well suited for practical applications such as continuous EEG monitoring, early seizure screening, and ambulatory ECG surveillance, where abnormal events are rare and annotations are often limited or unreliable.

Nevertheless, several limitations remain. First, although MAGE operates under an unsupervised anomaly detection paradigm, it still relies on predominantly normal training data,

and extending the framework to fully unsupervised or weakly supervised scenarios remains an important direction for future work. Second, while the current evaluation adopts a sample-level data split to ensure sufficient normal data diversity and stable training, true clinical generalization ideally requires reliable performance on unseen subjects. Achieving subject-wise generalization is particularly challenging for physiological datasets with limited subject sizes and high inter-subject variability, and this issue remains an open and critical problem that warrants further investigation.

Future work will therefore focus on improving cross-subject and cross-dataset generalization through more adaptive, subject-aware, and domain-invariant representation learning strategies. In addition, extending the framework to additional physiological modalities and further enhancing interpretability and clinical trustworthiness will be essential steps toward broader real-world deployment.

REFERENCES

- [1] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection," in *International conference on learning representations*, 2018.
- [2] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. v. d. Hengel, "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1705–1714.
- [3] M. Roy, S. Majumder, A. Halder, and U. Biswas, "Ecg-net: A deep lstm autoencoder for detecting anomalous ecg," *Engineering Applications of Artificial Intelligence*, vol. 124, p. 106484, 2023.
- [4] S. Dhyani and R. Butola, "Autoencoders for ecg anomaly detection: A survey," *Journal of Computing in Education, Sports and Health (JCESH)*, vol. 2, no. 1, pp. 47–58, 2025.
- [5] Y. Guan, W. Wang, J. Chen, P. Yang, J. Xu, and J. Qi, "A survey of multimodal fusion for alzheimer's disease prediction: A new taxonomy and trends," *Information Fusion*, p. 104098, 2025.
- [6] Y. Zhu, J. Jia, M. Dong, and J. Qi, "Sentirecache: Accelerate sentiment classification with saliency-based caching," *IEEE Transactions on Affective Computing*, vol. 16, no. 3, pp. 1349–1361, 2025.
- [7] N. Li, J. Tu, B. Yua, Z. Fu, S. Hua, Y. Guan, and J. Qi, "Mnsg-net: A mamba-optimized u-net for semantic gap reduction in lumbar disc herniation segmentation," *Biomedical Signal Processing and Control*, vol. 114, p. 109299, 2026.
- [8] J. Qi, R. Cai, Q. Liu, W. Wang, J. Ma, and J. Chen, "A dynamic bayesian multi-channel fusion scheme for heart rate monitoring with ballistocardiograph signals in free-living environments," *IEEE Journal of Selected Areas in Sensors*, 2024.
- [9] J. Gotman, "Automatic recognition of epileptic seizures in the eeg," *Electroencephalography and clinical Neurophysiology*, vol. 54, no. 5, pp. 530–540, 1982.

- [10] T. Gasser, P. Bäcker, and H. Steinberg, "Test-retest reliability of spectral parameters of the eeg," *Electroencephalography and clinical neurophysiology*, vol. 60, no. 4, pp. 312–319, 1985.
- [11] J. P. Martínez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, "A wavelet-based eeg delineator: evaluation on standard databases," *IEEE Transactions on biomedical engineering*, vol. 51, no. 4, pp. 570–581, 2004.
- [12] M. Zhou, C. Tian, R. Cao, B. Wang, Y. Niu, T. Hu, H. Guo, and J. Xiang, "Epileptic seizure detection based on eeg signals and cnn," *Frontiers in neuroinformatics*, vol. 12, p. 95, 2018.
- [13] W. Mao, H. Fathurrahman, Y. Lee, and T. Chang, "Eeg dataset classification using cnn method," in *Journal of physics: conference series*, vol. 1456, no. 1. IOP Publishing, 2020, p. 012017.
- [14] Z. Wang, L. Shi, H. Wu, J. Luo, X. Kong, and J. Qi, "Distilclip-eeg: Enhancing epileptic seizure detection through multi-modal learning and knowledge distillation," *IEEE Journal of Biomedical and Health Informatics*, 2025.
- [15] L. Shi, Z. Wang, Y. Ma, J. Chen, J. Xu, and J. Qi, "Double attention-based deep convolutional neural network for seizure detection using eeg signals," in *International Conference on Intelligent Computing*. Springer, 2024, pp. 392–404.
- [16] S. Roy, I. Kiral-Kornek, and S. Harrer, "Chrononet: A deep recurrent neural network for abnormal eeg identification," in *Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings 17*. Springer, 2019, pp. 47–56.
- [17] S. Kuila, N. Dhandu, and S. Joardar, "Ecg signal classification and arrhythmia detection using elm-rnn," *Multimedia Tools and Applications*, vol. 81, no. 18, pp. 25 233–25 249, 2022.
- [18] D. P. Kingma, M. Welling *et al.*, "Auto-encoding variational bayes," 2013.
- [19] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," *arXiv preprint arXiv:1511.05644*, 2015.
- [20] R. Daci, A. Taleb-Ahmed, L. Patrono, and C. Distant, "Sparse temporal autoencoder for eeg anomaly detection," *Sensors*, vol. 26, no. 5, p. 1589, 2026.
- [21] S. Sukhbaatar, J. Weston, R. Fergus *et al.*, "End-to-end memory networks," *Advances in neural information processing systems*, vol. 28, 2015.
- [22] C. Zhang, Y. Wang, X. Zhao, Y. Guo, G. Xie, C. Lv, and B. Lv, "Memory-augmented anomaly generative adversarial network for retinal oct images screening," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1971–1974.
- [23] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. v. d. Hengel, "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1705–1714.
- [24] M. Yang, Y. Li, Z. Huang, Z. Liu, P. Hu, and X. Peng, "Partially view-aligned representation learning with noise-robust contrastive loss," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 1134–1143.
- [25] T. Xiao, Z. Wang, Y. Zhang, S. Wang, H. Feng, Y. Zhao *et al.*, "Self-supervised learning with attention mechanism for eeg-based seizure detection," *Biomedical Signal Processing and Control*, vol. 87, p. 105464, 2024.
- [26] P. Chen, M. Deng, D. Zhang, B. Zhou, and F. Haoyi, "Noise-aware self-supervised denoising for electrocardiogram anomaly detection," Available at SSRN 4601063.
- [27] H. Han, H. Fan, X. Huang, and C. Han, "Self-supervised multi-transformation learning for time series anomaly detection," *Expert Systems with Applications*, vol. 253, p. 124339, 2024.
- [28] G. Bao, B. Yan, L. Tong, J. Shu, L. Wang, K. Yang, and Y. Zeng, "Data augmentation for eeg-based emotion recognition using generative adversarial networks," *Frontiers in Computational Neuroscience*, vol. 15, p. 723843, 2021.
- [29] N. E. H. S. B. Aissa, C. A. Kerrache, A. Korichi, A. Lakas, and A. N. Belkacem, "Enhancing eeg signal classifier robustness against adversarial attacks using a generative adversarial network approach," *IEEE Internet of Things Magazine*, vol. 7, no. 3, pp. 44–49, 2024.
- [30] H. Yi, Y. Ming, D. Liu, and W. Kong, "Time-frequency jointed imperceptible adversarial attack to brainprint recognition with deep learning models," in *2024 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2024, pp. 1–6.
- [31] I. Obeid and J. Picone, "The temple university hospital eeg data corpus," *Frontiers in Neuroscience*, vol. 10, 2016. [Online]. Available: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2016.00196>
- [32] V. Shah, E. von Weltin, S. Lopez, J. R. McHugh, L. Veloso, M. Golmohammadi, I. Obeid, and J. Picone, "The temple university hospital seizure detection corpus," *Frontiers in Neuroinformatics*, vol. 12, 2018. [Online]. Available: <https://www.frontiersin.org/journals/neuroinformatics/articles/10.3389/fninf.2018.00083>
- [33] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.
- [34] J. Gutttag. (2010) Chb-mit scalp eeg database (version 1.0.0). PhysioNet. [Online]. Available: <https://doi.org/10.13026/C2K01R>
- [35] W. Nasreddine, "Epileptic eeg dataset. 2021," URL <https://doi.org/10.17632/5pc2j46cbc>, vol. 1, 2021.
- [36] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [37] R. Bousseltot, D. Kreiseler, and A. Schnabel, "Nutzung der ekg-signal-datenbank cardiodat der ptb über das internet," 1995.
- [38] Y. Song, Q. Zheng, B. Liu, and X. Gao, "Eeg conformer: Convolutional transformer for eeg decoding and visualization," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 710–719, 2023.
- [39] D. Priyasad, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Interpretable seizure classification using unprocessed eeg with multi-channel attentive feature fusion," *IEEE Sensors Journal*, vol. 21, no. 17, pp. 19 186–19 197, 2021.
- [40] J. Yin, A. Liu, C. Li, R. Qian, and X. Chen, "A gan guided parallel cnn and transformer network for eeg denoising," *IEEE Journal of Biomedical and Health Informatics*, 2023.
- [41] C.-S. Tung, S.-F. Liang, S.-F. Chang, and C.-P. Young, "A hybrid artificial intelligence system for automated eeg background analysis and report generation," *IEEE Journal of Biomedical and Health Informatics*, 2024.
- [42] G. U. Nneji, M. A. Hossin, K. D. Mark, E. S. Umana, G. T. Mgbejime, and J. Li, "Enhancing eeg classification in cardiac diagnostics: A novel approach using adaptive focal cross-entropy loss function," *IEEE Journal of Biomedical and Health Informatics*, 2025.
- [43] T.-Y. Kim and S.-B. Cho, "Web traffic anomaly detection using c-lstm neural networks," *Expert Systems with Applications*, vol. 106, pp. 66–76, 2018.
- [44] S. Roy, I. Kiral-Kornek, and S. Harrer, "Chrononet: A deep recurrent neural network for abnormal eeg identification," in *Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings 17*. Springer, 2019, pp. 47–56.
- [45] A. Ortiz, F. J. Martinez-Murcia, J. L. Luque, A. Giménez, R. Morales-Ortega, and J. Ortega, "Dyslexia diagnosis by eeg temporal and spectral descriptors: an anomaly detection approach," *International Journal of Neural Systems*, vol. 30, no. 07, p. 2050029, 2020.
- [46] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga, "Usad: Unsupervised anomaly detection on multivariate time series," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 3395–3404.
- [47] Y. Zhang, Y. Chen, J. Wang, and Z. Pan, "Unsupervised deep anomaly detection for multi-sensor time-series signals," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 2, pp. 2118–2132, 2021.
- [48] J. Xu, H. Wu, J. Wang, and M. Long, "Anomaly transformer: Time series anomaly detection with association discrepancy," *arXiv preprint arXiv:2110.02642*, 2021.
- [49] S. Liu, B. Zhou, Q. Ding, B. Hooi, Z. Zhang, H. Shen, and X. Cheng, "Time series anomaly detection with adversarial reconstruction networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 4293–4306, 2022.
- [50] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "Timesnet: Temporal 2d-variation modeling for general time series analysis," *arXiv preprint arXiv:2210.02186*, 2022.
- [51] S. Rukhsar and A. K. Tiwari, "Lightweight convolution transformer for cross-patient seizure detection in multi-channel eeg signals," *Computer Methods and Programs in Biomedicine*, vol. 242, p. 107856, 2023.
- [52] J. Duan, F. Xie, N. Huang, N. Luo, Z. Guan, W. Zhao, and G. Gao, "An eeg abnormality detection algorithm based on graphic attention network," *Multimedia Tools and Applications*, vol. 83, no. 6, pp. 17 941–17 960, 2024.

- [53] Y. Zhao and K. Iramina, "Cae-t: A channelwise autoencoder with transformer for eeg abnormality detection," *arXiv preprint arXiv:2412.14522*, 2024.
- [54] Y. Wang, S. Yuan, J.-X. Liu, W. Hu, Q. Jia, and F. Xu, "Combining eeg features and convolutional autoencoder for neonatal seizure detection," *International Journal of Neural Systems*, vol. 34, no. 08, p. 2450040, 2024.
- [55] J. Wang, M. Sun, and W. Huang, "Unsupervised eeg-based seizure anomaly detection with denoising diffusion probabilistic models," *International Journal of Neural Systems*, vol. 34, no. 09, p. 2450047, 2024.
- [56] B. Zheng, L. Ming, K. Zeng, M. Zhou, X. Zhang, T. Ye, B. Yang, X. Zhou, and C. S. Jensen, "Adversarial graph neural network for multivariate time series anomaly detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 12, pp. 7612–7626, 2024.
- [57] S. Chen, D. Huang, X. Liu, J. Chen, X. Kong, and T. Zhang, "Automatic diagnostics of electroencephalography pathology based on multi-domain feature fusion," *PLoS One*, vol. 20, no. 5, p. e0310348, 2025.
- [58] I. Naidji, A. Tibermacine, I. E. Tibermacine, S. Russo, and C. Napoli, "Egdn-kl: Dynamic graph-deviation network for eeg anomaly detection," *Biomedical Signal Processing and Control*, vol. 112, p. 108597, 2026.
- [59] H. Shi and Z. Jiang, "Lavae-net: A deep lstm-attention vae framework for detecting ecg anomalies," *Biomedical Signal Processing and Control*, vol. 118, p. 109755, 2026.
- [60] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.
- [61] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," *arXiv preprint arXiv:1706.06083*, 2017.
- [62] N. Carlini and D. Wagner, "Towards evaluating the robustness of neural networks," in *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2017, pp. 39–57.
- [63] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. v. d. Hengel, "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.